



# Mining the Bilinear Structure of Data with Approximate Joint Diagonalization

Louis Korczowski, Florent Bouchard, Christian Jutten, Marco Congedo

## ► To cite this version:

Louis Korczowski, Florent Bouchard, Christian Jutten, Marco Congedo. Mining the Bilinear Structure of Data with Approximate Joint Diagonalization. EUSIPCO 2016 - 24th European Signal Processing Conference, EURASIP, Aug 2016, Budapest, Hungary. pp.667-671. hal-01357245

**HAL Id: hal-01357245**

**<https://hal.science/hal-01357245>**

Submitted on 29 Aug 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Mining the Bilinear Structure of Data with Approximate Joint Diagonalization

Louis Korczowski, Florent Bouchard, Christian Jutten, Marco Congedo  
GIPSA-lab, Univ. Grenoble Alpes, CNRS, Grenoble Institute of Technology, Grenoble France  
Email: louis.korczowski@gipsa-lab.fr

**Abstract**—Approximate Joint Diagonalization of a matrix set can solve the linear Blind Source Separation problem. If the data possesses a bilinear structure, for example a spatio-temporal structure, transformations such as tensor decomposition can be applied. In this paper we show how the linear and bilinear joint diagonalization can be applied for extracting sources according to a composite model where some of the sources have a linear structure and other a bilinear structure. This is the case of Event Related Potentials (ERPs). The proposed model achieves higher performance in term of shape and robustness for the estimation of ERP sources in a Brain Computer Interface experiment.

## I. INTRODUCTION

Event-Related Potentials (ERPs) are transient brainwaves with a fixed *evoked* spatio-temporal pattern ; their polarity, shape, latency and amplitude are approximatively constant across observations for a given class of stimuli evoking them. The amplitude of ERPs is known to be very low compared to the ongoing *continuous* background activity of the brain. For many applications such as classification or clinical analysis, one would like to retrieve both *evoked* and *continuous* sources of the brain (Figure 1).

Blind Source Separation (BSS) is a fundamental framework for modeling independent sources hidden in the observations. BSS methods find major applications in telecommunication, biomedical engineering, speech, audio and video processing [1].

Let us consider the instantaneous linear mixture given by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where  $t$  is the time index,  $\mathbf{x}(t), \mathbf{s}(t) \in \mathbb{R}^N$  are respectively the *observation* and *source* vectors,  $\mathbf{A} \in \mathbb{R}^{N \times N}$  is an invertible unknown *spatial mixing matrix*, and  $\mathbf{n}(t) \in \mathbb{R}^N$  is an additive *noise* vector. In model (1) the fixed pattern is only spatial. We can formulate the problem of estimating  $\mathbf{s}(t)$  by finding a *spatial unmixing matrix*  $\mathbf{B}$  that solves  $\hat{\mathbf{s}}(t) = \mathbf{B}^T \mathbf{x}(t)$  where the sources  $\hat{\mathbf{s}} = \mathbf{P}\Delta\mathbf{s}$  are estimated up to a permutation matrix  $\mathbf{P}$  and a diagonal matrix (scaling)  $\Delta$  ambiguity ; superscript  $(.)^T$  is the transpose operator.

Approximate Joint Diagonalization (AJD) is a class of methods for solving the BSS problem relying on the diagonalization of a set of symmetric matrices  $\mathbf{R}_x(l) \forall l \in \{1, \dots, L\}$ , named *target* matrices, containing statistics of the observation  $\mathbf{x}$ . The choice of the statistics (e.g. correlations, time-lagged covariances, co-spectra, higher order cumulants, etc.) depends on assumptions on the sources such as non-stationarity, spectral coloration, non-Gaussianity, etc. [1], [2]. Indices  $l$  can stand,

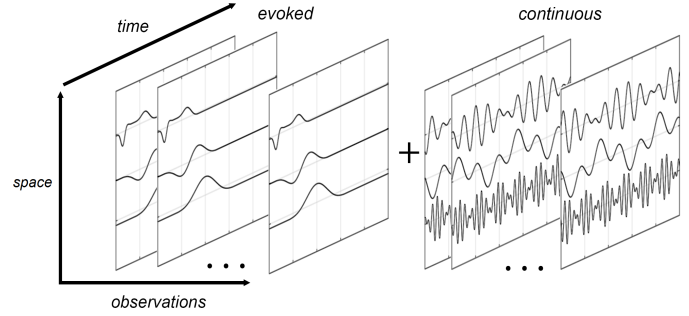


Fig. 1. The *composite* model assumes that the observations are the superposition of evoked sources and continuous sources linked by the same spatial mixing matrix  $\mathbf{A}$ .

e.g., for delays  $\tau$  in the case of covariances or for frequencies  $f$  in the case of co-spectra, etc. [2]. Thus, the BSS problem can be reformulated as

$$\mathbf{R}_x(l) = \mathbf{A}\mathbf{R}_s(l)\mathbf{A}^T, \quad (2)$$

where  $\mathbf{R}_s(l)$  is a diagonal matrix containing the power of the sources. We seek a matrix  $\mathbf{B}$  that jointly diagonalizes all matrices  $\mathbf{R}_x(l)$  in (2) *as much as possible* (i.e., according to some criterion)

$$\mathbf{R}_{\hat{s}}(l) = \mathbf{B}^T \mathbf{R}_x(l) \mathbf{B}. \quad (3)$$

AJD has been successfully used in source separation of *continuous* electroencephalographic (EEG) sources [3] using co-spectral as *target* matrices. More recently, [4] proposed to add the covariance matrices of the estimated *evoked* activity for the separation of Event Error Related Potential (ErrP) sources. In ERPs, the *evoked* sources have both a spatial and a temporal (bilinear) fixed structure such as

$$\mathbf{X}_k = \mathbf{A}\mathbf{S}_k\mathbf{E}^T + \mathbf{N}_k, \quad (4)$$

where matrices  $\mathbf{X}_k \in \mathbb{R}^{N \times T}$  are a set of observation matrices with  $k \in \{1, \dots, K\}$ ,  $\mathbf{E} \in \mathbb{R}^{T \times N}$  is an unknown *temporal mixing matrix* and  $\mathbf{S}_k \in \mathbb{R}^{N \times N}$  is a diagonal matrix containing the amplitude of the sources and  $\mathbf{N}_k \in \mathbb{R}^{N \times T}$  is a noise term. The difference with the linear BSS case (1) is that the bilinear model has now also a fixed temporal structure, contained in  $\mathbf{E}$ , which can be modulated in amplitude across observations by  $\mathbf{S}_k$ . It is a bilinear BSS problem with

$$\hat{\mathbf{S}}_k = \mathbf{B}^T \mathbf{X}_k \mathbf{D}, \quad (5)$$

where now two unmixing matrices  $\mathbf{B}$  and  $\mathbf{D}$  are required for diagonalizing *as much as possible*  $\hat{\mathbf{S}}_k$ . A proposed algorithm solved it by Approximate Joint Singular Value Decomposition (AJSVD) in [5] considering  $\mathbf{A}$  and  $\mathbf{E}$  orthogonal. More generally, the model can be solved by tensor rank decomposition, i.e. Candecomp/PARAFAC (CP) [6].

Such model was used for the extraction of the ERP sources in functional Magnetic Resonance Imaging (fMRI) [7] and a solution using AJD was proposed in [8] by diagonalizing both spatial and temporal covariance matrices. Recently, CP decomposition was used in addition to Extended Kalman Filter in [9] with successful estimation of the electroencephalographic ERP at single trial level.

However, in real data such as ERP, the observation  $\mathbf{X}_k$  is an additive process (see Figure 1) of both *continuous* activity which is separable by the linear model (3), and *evoked* activity, which is separable by the bilinear model (5). Thus trying a linear BSS (respectively bilinear BSS) method to extract the *continuous* (respectively *evoked*) sources will fail because the model neglects the contribution of the *evoked* (respectively *continuous*) sources. Eventually none of these models can find the *true* unmixing matrices  $\mathbf{B}, \mathbf{D}$ .

To go beyond those limitations, we propose in Section II-A a composite model that can be used to solves simultaneously the AJD for (1) and a bilinear AJD (BAJD) for (4). We name the proposed model *composite AJD* (CAJD). We give the corresponding cost functions for AJD, BAJD and CAJD in Section II-B and a closed-form minimizer of the respective cost function by Gauss Planar Transformation (GTP) in Section II-C, using yet a novel Jacobi-like algorithm. We compare the proposed models in the estimation of matrices  $\mathbf{B}$  and  $\mathbf{E}$  on simulated data in Section III-A and we compare BAJD vs. CAJD on EEG ERP in Section III-C.

## II. METHODS

### A. Composite Approximate Joint Diagonalization

One would like to complete the full source separation by jointly diagonalize  $\mathbf{B}^T \mathbf{R}_x(l) \mathbf{B}$  and  $\mathbf{B}^T \mathbf{X}_k \mathbf{D}$  where  $\mathbf{B}$  and  $\mathbf{D}$  are respectively the spatial and temporal unmixing matrices supposed invertible. Note that if  $\mathbf{E}$  in model (4) is furthermore supposed orthogonal and its estimation is not of interest for the user, one can always consider the classical AJD framework as  $\mathbf{X}_k \mathbf{X}_k^T = \mathbf{A} \mathbf{S}_k \mathbf{E} \mathbf{E}^T \mathbf{S}_k \mathbf{A}^T = \mathbf{A} \mathbf{R}_s(k) \mathbf{A}^T \forall k \in \{1, \dots, K\}$  (with  $\mathbf{S}_k$  diagonal), coming back to (2).

### B. Cost Functions

In the linear model case given by (3) and using the standard AJD framework, we consider the cost function  $f$ ,

$$f(\mathbf{B}) = \sum_{l=1}^L \left\| \text{off}(\mathbf{B}^T \mathbf{R}_x(l) \mathbf{B}) \right\|_F^2, \quad (6)$$

where matrices  $\mathbf{R}_x(l)$  correspond to the chosen statistics of  $\mathbf{X}_k$  and  $\text{off}(\cdot)$  is the off-diagonal operator that set all diagonal elements of the argument to zero.

For the bilinear model (4), BAJD uses a similar cost function  $f_b$ , where  $b$  stand for *bilinear*, given by

$$f_b(\mathbf{B}, \mathbf{D}) = \sum_{k=1}^K \left\| \text{off}(\mathbf{B}^T \mathbf{X}_k \mathbf{D}) \right\|_F^2, \quad (7)$$

where matrices  $\mathbf{X}_k$  correspond to observations in spatio-temporal domain (i.e., for ERP, the raw potential and not their covariance) and  $\mathbf{B}$  and  $\mathbf{D}$  are their corresponding spatial and temporal unmixing matrices, respectively.

Finally for the composite model introduced in the previous section, the corresponding cost function  $f_c$ , where  $c$  stand for *composite*, is given by

$$f_c(\mathbf{B}, \mathbf{D}) = \alpha f(\mathbf{B}) + (1 - \alpha) f_b(\mathbf{B}, \mathbf{D}), \quad (8)$$

where  $\alpha \in [0, 1]$  is a fixed ponderation weight which expresses a preference among the two models if  $\alpha \neq 0.5$ , i.e linear ( $> 0.5$ ) or bilinear ( $< 0.5$ ). In this work, we consider  $\alpha = 0.5$ .

### C. Optimization by Gauss Planar Transformations

To solve the problem for the three functions we consider, we propose a novel optimization scheme inspired by eigenvector Jacobi and elementary Gauss elimination methods, named Gauss Planar Transformation (GPT). Note that we choose the same optimization scheme for all methods in order to compare the models fairly. Starting from initial guess matrices  $\mathbf{B}$  and  $\mathbf{D}$  (see Section II-D), we perform repeated planar non-orthogonal transformations on the columns of the unmixing matrices. Given  $i$  and  $j$  in  $\{1, \dots, N\}$  with  $i \neq j$ , a  $(i, j)$ -transformation is defined as

$$\begin{aligned} \mathbf{b}_i &\leftarrow \mathbf{b}_i + \beta \mathbf{b}_j \\ \mathbf{d}_i &\leftarrow \mathbf{d}_i + \gamma \mathbf{d}_j \end{aligned}$$

where  $\mathbf{b}_i$  (respectively  $\mathbf{d}_i$ ) denotes the  $i^{\text{th}}$  column of  $\mathbf{B}$  (respectively  $\mathbf{D}$ ). A sweep of the algorithms is then to perform  $(i, j)$ -transformations for all  $i$  and  $j$  in  $\{1, \dots, N\}$  with  $i \neq j$ . We seek parameters  $\beta$  and  $\gamma$  that minimize the cost function for each  $(i, j)$ -transformation. Such a transformation acts only on the row  $i$  and the column  $i$  of the matrices  $\mathbf{R}_x(l)$  and/or  $\mathbf{X}_k$ . A solution would then to find  $\beta$  and  $\gamma$  in order to minimize all those elements (except the diagonal one) as it is done in [10]. Another way is to only minimize the elements  $(i, j)$  and  $(j, i)$  as in [11]. We choose the latter approach as this solution always converges faster and does not affect the performance (data not shown). Given  $i$  and  $j$ , this leads, in the AJD case (3), to minimize

$$f^{ij}(\beta) = 2 \sum_{l=1}^L ((\mathbf{b}_i^T + \beta \mathbf{b}_j^T) \mathbf{R}_x(l) \mathbf{b}_j)^2$$

For the BAJD model in (5), we have to minimize

$$\begin{cases} f_b^{ij}(\beta) = \sum_{k=1}^K ((\mathbf{b}_i^T + \beta \mathbf{b}_j^T) \mathbf{X}_k \mathbf{d}_j)^2 \\ f_b^{ij}(\gamma) = \sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k (\mathbf{d}_i + \gamma \mathbf{d}_j))^2. \end{cases}$$

Finally, if we consider the CAJD model proposed in Section II-A, the functionals to minimize are

$$\begin{cases} f_c^{ij}(\beta) &= (1-\alpha) \sum_{k=1}^K ((\mathbf{b}_i^T + \beta \mathbf{b}_j^T) \mathbf{X}_k \mathbf{d}_j)^2 \\ &+ 2\alpha \sum_{f=1}^F ((\mathbf{b}_i^T + \beta \mathbf{b}_j^T) \mathbf{R}_x(l) \mathbf{b}_j)^2 \\ f_c^{ij}(\gamma) &= (1-\alpha) \sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k (\mathbf{d}_i + \gamma \mathbf{d}_j))^2 \end{cases}.$$

All those functions are second order polynomials with closed form minimizers. Thus, for AJD with cost function (6), the optimal  $\beta$  is given by

$$\beta = \frac{\sum_{k=1}^K (\mathbf{b}_i^T \mathbf{R}_x(l) \mathbf{b}_j) (\mathbf{b}_j^T \mathbf{R}_x(l) \mathbf{b}_j)}{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{R}_x(l) \mathbf{b}_j)^2}.$$

For the BAJD model with cost function (7), optimal  $\beta$  and  $\gamma$  are given by

$$\begin{aligned} \beta &= \frac{\sum_{k=1}^K (\mathbf{b}_i^T \mathbf{X}_k \mathbf{d}_j) (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)}{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)^2} \\ \gamma &= \frac{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_i) (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)}{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)^2}. \end{aligned}$$

Finally, for our composite model CAJD with the cost function (8), optimal  $\beta$  and  $\gamma$  are given by

$$\begin{aligned} \beta &= \frac{(1-\alpha) \sum_{k=1}^K (\mathbf{b}_i^T \mathbf{X}_k \mathbf{d}_j) (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j) + 2\alpha \sum_{f=1}^F (\mathbf{b}_i^T \mathbf{R}_x(l) \mathbf{b}_j) (\mathbf{b}_j^T \mathbf{R}_x(l) \mathbf{b}_j)}{(1-\alpha) \sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)^2 + 2\alpha \sum_{f=1}^F (\mathbf{b}_j^T \mathbf{R}_x(l) \mathbf{b}_j)^2} \\ \gamma &= \frac{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_i) (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)}{\sum_{k=1}^K (\mathbf{b}_j^T \mathbf{X}_k \mathbf{d}_j)^2}. \end{aligned}$$

#### D. Initialization

The initialization is of paramount importance for Jacobi-like algorithms. While a working initialization can be the identity matrix, we use singular value decomposition:

$$\bar{\mathbf{X}}^T = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \quad (9)$$

where  $\bar{\mathbf{X}}$  is an estimation of the evoked activity such as the arithmetic average  $\bar{\mathbf{X}} = \sum_k^K \frac{1}{K} \mathbf{X}_k$  or by regression based estimation [12]. We set

$$\begin{cases} \mathbf{B}_0 = \mathbf{U} \\ \mathbf{D}_0 = \tilde{\mathbf{V}}^T \end{cases} \quad (10)$$

where  $\tilde{\mathbf{V}} \in \mathbb{R}^{T \times N}$  is the reduced right-handed component of the SVD-decomposition containing only the  $N$  singular vectors with the largest  $N$  associated singular values in  $\mathbf{\Lambda}$ .

### III. RESULTS

#### A. Simulations

In order to estimate the performances of CAJD compared to AJD and BAJD on the model that we consider, we first performed simulations within Matlab (c) environnement.

To analyze the quality of the spatial and temporal unmixing matrices found by both methods, we use the Moreau-Macchi index  $I_{M-M}$  [13] defined as

$$I_{M-M} = \frac{1}{2(n-1)} \left( \sum_{i=1}^n \left[ \frac{\sum_{j=1}^n |H_{ij}|}{\max_{1 \leq j \leq n} |H_{ij}|} - 1 \right] + \sum_{j=1}^n \left[ \frac{\sum_{i=1}^n |H_{ij}|}{\max_{1 \leq i \leq n} |H_{ij}|} - 1 \right] \right), \quad (11)$$

where  $\mathbf{H}$  equals  $\mathbf{B}^T \mathbf{A}$  (respectively  $\mathbf{E}^T \mathbf{D}$ ), with  $\mathbf{B}$  (respectively  $\mathbf{D}$ ) the estimated spatial (respectively temporal) unmixing matrix and  $\mathbf{A}$  (respectively  $\mathbf{E}$ ) the true spatial (respectively temporal) mixing matrix.

We generated  $K = 100$  matrices  $\mathbf{X}_k$  of dimension  $N \times T$  with  $N = 16$  and  $T = 128$  according to

$$\mathbf{X}_k = \mathbf{A} \mathbf{S}_k \mathbf{E}^T + \sigma \mathbf{N}_k, \quad (12)$$

where  $\mathbf{A}$  is the spatial mixing matrix being non-orthogonal generated randomly with i.i.d. elements normally distributed  $\mathcal{N}(0, 1)$ .  $\mathbf{E}$  is the temporal mixing matrix with i.i.d. elements generated from a multivariate normal distribution with a random covariance matrix. We controlled the non-orthogonality of  $\mathbf{A}$  and  $\mathbf{E}$  by constraining their condition number with respect to inversion between 1 and 20 (excluded). Matrices  $\mathbf{S}_k$  are diagonal with i.i.d. elements randomly drawn from a normal distribution  $\mathcal{N}(0, 1)$ .  $\sigma$  is a free parameter defining the noise level and  $\mathbf{N}_k$  are noise matrices with elements drawn independently from a normal distribution  $\mathcal{N}(0, 1)$ .

In parallel, we generated  $L = 100$  matrices  $\mathbf{R}_x(l)$  of dimension  $N \times N$  according to

$$\mathbf{R}_x(l) = \mathbf{A} \mathbf{R}_s(l) \mathbf{A}^T + \frac{\sigma^2}{2} (\mathbf{N}_l + \mathbf{N}_l^T),$$

where  $\mathbf{A}$  and  $\sigma$  are the same as before. Matrices  $\mathbf{R}_s(l)$  are diagonal with i.i.d. elements generated from a chi-squared distribution with two degrees of freedom. The symetrized noise matrices  $\mathbf{N}_l$  are with elements drawn independently from a normal distribution  $\mathcal{N}(0, 1)$ .

AJD was performed on the matrices  $\mathbf{X}_k \mathbf{X}_k^T$  and  $\mathbf{R}_x(l)$

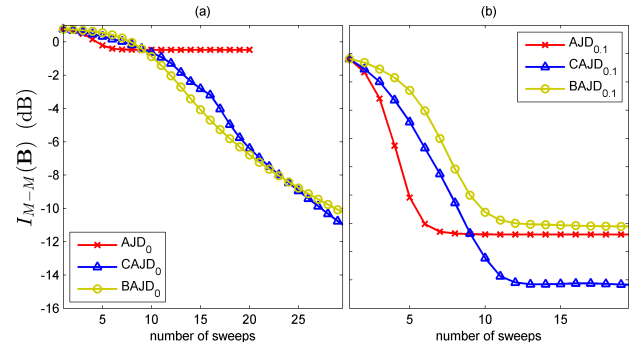


Fig. 2. Moreau-Macchi criteria on the matrix  $\mathbf{B}$ . Average performance of GPT algorithms on 100 independant realizations (a) without noise ( $\sigma = 0$ ) and (b) with noise ( $\sigma = 0.1$ ).

TABLE I  
MOREAU-MACCHI CRITERION AFTER CONVERGENCE. 100 INDEPENDENT  
DRAWS OF MATRICES  $\mathbf{A}$  AND  $\mathbf{E}$  WITH  $\sigma = 0.1$ .

		median	10% quantile	90% quantile
$\mathbf{B}$	CAJD	$1.06e-2$	$4.79e-2$	$2.69e-2$
	BAJD	$3.12e-1$	$1.21e-1$	$7.09e-1$
	AJD	$3.26e-1$	$2.87e-1$	$3.78e-1$
$\mathbf{D}$	CAJD	$4.26e-1$	$1.24e-1$	$8.74e-1$
	BAJD	$4.17e-1$	$1.28e-1$	$8.81e-1$

and  $\mathbf{B}$  was initialized with the identity matrix. BAJD was performed on the matrices  $\mathbf{X}_k$ .  $\mathbf{B}$  was initialized in the same way as AJD and  $\mathbf{D}$  with an orthogonalized random matrix. CAJD was performed on both the matrices  $\mathbf{X}_k$  and  $\mathbf{R}_x(l)$  with the same initialization as BAJD.

As expected, because  $\mathbf{E}$  is far from being orthogonal, AJD gives results far from the true solution, see Figure 2. Moreover, while we can not differentiate the performance of BAJD and CAJD in the case of no noise, the convergence is always better for CAJD in presence of noise for the estimation of  $\mathbf{B}$  (see Table I). The difference in the estimation of  $\mathbf{E}$  is not significant between BAJD and CAJD because the simulated data respect the bilinear model.

Thus, CAJD displays overall better performances as compared to both AJD and BAJD on the estimation of  $\mathbf{B}$ . Our model is equivalent to BAJD for the estimation of  $\mathbf{D}$ . Next, we will test both BAJD and CAJD on real data.

### B. Procedures

1) *Data Acquisition*: EEG signals were acquired during sessions of two subjects playing simultaneously to a P300-based BCI [14]. EEG data were acquired with 32 active wet electrodes using USBamp amplifiers [g.Tec, Graz, Austria] with a sample rate  $F_s=512$  Hz. Ground was an active electrode on Fz and reference was a passive electrode clip on the right earlobe. The ERP triggers were synchronized with the signals with a supplementary analog channel to minimize the jitter.

2) *Preprocessing*: From the recorded signals, we selected 16 representative electrodes from the first subject ( $N=16$ ): Fp1-Fp2-F5-AFz-F6-T7-Cz-T8-P7-P3-Pz-P4-P8-O1-Oz-O2. The signals were filtered by a fourth order forward-backward Butterworth band pass filter [1-20] Hz and down-sampled at  $F_s=128$ Hz. The signals were segmented into trials  $\mathbf{X}_k$  of 1s ( $T=128$ ) starting at each visual stimulation. Each trial was labeled as TARGET (TA), i.e. with possible appearance of P300 potential, or NON-TARGET (NT) for the two different experimental conditions.

3) *Target Matrices*: We propose to estimate the *evoked* activity  $\bar{\mathbf{X}}_z$  by arithmetic ensemble average of  $\mathbf{X}_k$  by bootstrapping 50 new ensembles: 40 ensemble averages of 50 observations in the condition TA; 10 ensemble averages of 50 observations in the condition NT.

Thus for AJD (6), we propose target matrices  $\mathbf{R}_x(l) \in \{\mathbf{C}_{\bar{\mathbf{X}}_z}, \mathbf{C}_f\}$  where  $\mathbf{C}_f$  are co-spectral matrices computed by Bartlett's average [15] over the  $K$  observations for frequencies

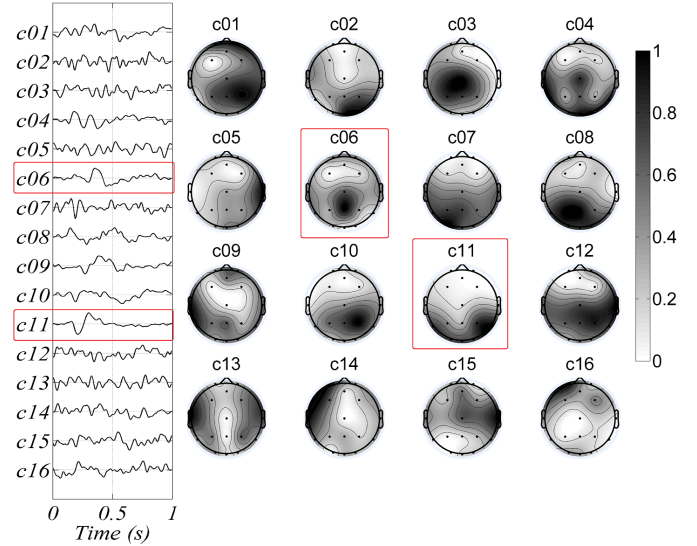


Fig. 3. Left: columns of  $\hat{\mathbf{E}}_i$  normalized with L2-norm. Right: columns of  $\hat{\mathbf{A}}_i$  projected on a topographic scalp map in absolute amplitude and normalized with . Highlight : Source c06 corresponds to P3b, source c11 corresponds to early visual complex.

of interest, i.e.  $f \in \{1, 1.5, 2, \dots, 20\} \text{ Hz}$  for a total of 39 co-spectral matrices.  $\mathbf{C}_{\bar{\mathbf{X}}_z}$  are the empirical covariance matrices of the  $\bar{\mathbf{X}}_z \forall z \in \{1, \dots, 50\}$ . For BAJD (7), we diagonalized  $\bar{\mathbf{X}}_z$ . Eventually, for CAJD we replace  $\mathbf{X}_k$  by  $\bar{\mathbf{X}}_z$  and  $\mathbf{R}_x(l)$  by  $\mathbf{C}_f$  in (8).

4) *GPT initialization*: The initialization of  $\mathbf{B}$ ,  $\mathbf{D}$  was made by a SVD step such as proposed in Section II-D with  $\bar{\mathbf{X}}$  estimated with the method in [12] (no subspace reduction).

### C. ERP Source Estimation

1) *Source Separation*: To assess the quality of the source separation obtained by CAJD, we compared the estimated sources according to their spatial and temporal contributions. We projected the columns of  $\hat{\mathbf{A}} = \mathbf{B}^{-1}$  on a scalp topographic map. The temporal source pattern is given by the columns  $\hat{\mathbf{E}} = \mathbf{D}^+$  where superscript  $(.)^+$  denotes the Moore-Penrose pseudoinverse.

In Figure 3, we present the resulting estimation obtained by CAJD on a representative subject of our dataset. By visual inspection, we can observe that several sources correspond to *evoked* potential: a) c11, the early visual complex, generated in the occipital (visual) cortex. This is common to conditions TA and NT but with a much lower amplitude in NT. b) c06 the P3b component of the ERP generated in parietal locations and with maximum amplitude around 300ms after the stimulus [16].

2) *Source Quality*: We focus on the quality of the estimated source associated to the *evoked* activity P3b of BAJD and CAJD in condition TA. Figure 4 displays their respective estimation (described in the following) at the single trial level and they are compared to the observation  $\mathbf{X}_k$ .



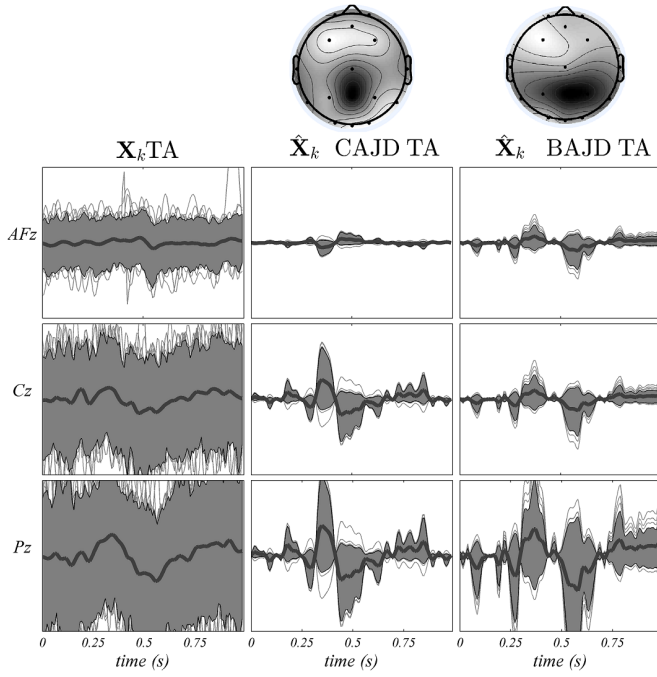


Fig. 4. TA condition: Data  $\mathbf{X}_k$  (left) and backprojected P3b source (13) for CAJD (center) and BAJD (right). Their respective spatial distribution is shown on top. y-axis are in relative amplitude. Thin grey line: observation ( $K=100$ ). Thick black line: Ensemble average. Grey area: 10% and 90% quantiles, greater is the area, larger is the variability.

For a given observation  $\mathbf{X}_k$ , we estimated the corresponding power of the sources in  $\hat{\mathbf{S}}_k$  and backprojected the  $i$ -th source in the electrode domain such as:

$$\hat{\mathbf{X}}_{k,i} = \hat{\mathbf{A}}_i \hat{\mathbf{S}}_k \hat{\mathbf{E}}_i^T \quad (13)$$

where  $\hat{\mathbf{X}}_{k,i}$  is the projection of the  $i$ -th component in the sensor domain.  $\hat{\mathbf{A}}_i$  and  $\hat{\mathbf{E}}_i$  are respectively the spatial and temporal contributions of the  $i$ -th estimated component. The projected activity of the source P3b in Figure 4 shows that the sign and shape of the source is properly estimated for both methods. The polarity of the estimated P3b of CAJD is constant over condition TA while it is not for condition NT (data not shown here). Interestingly, the CAJD is more robust to outliers as compared to BAJD, particularly in the early and late latency (before 250ms and after 700ms). Under this evidence, CAJD seems to estimate much better the real (physiological) source of the P3b.

#### IV. CONCLUSION

We have described a new approach for extracting event-related sources by mining the bilinear (spatio-temporal) structure of the data with a composite model of both AJD and BAJD named Composite Approximate Joint Diagonalization (CAJD). We proposed to minimize the off-diagonal criteria by a novel algorithm, i.e. Gauss Planar Transformation (GPT). We provided the close-form approximate minimizer for the three models in the framework of GPT. CAJD achieves better performance than AJD and BAJD models on simulated data.

Moreover this method is relevant for single trial source extraction in the case of Event-Related Potentials. We believe that CAJD could be applied on various situations such as denoising or classification for Brain Computer Interfaces. Eventually, the CAJD is not restricted to brain data but can be used for all combination of linear and bilinear structure whenever their relation is known.

#### ACKNOWLEDGMENT

This research is partly supported by the ERC CHES (2012-ERC-AdG-320684). F. Bouchard is partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01). We also thanks A. Andreev, engineer in charge of the software platform, and E. Ostaschenko, M. Cederhout for the experimental support.

#### REFERENCES

- [1] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, February 2010.
- [2] F.J. Theis and Y. Inouye. On the use of joint diagonalization in blind signal processing. In *2006 IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings*, pages 4 pp.–3589, 2006.
- [3] Marco Congedo, Cdric Gouy-Pailler, and Christian Jutten. On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, 119(12):2677–2686, December 2008.
- [4] Marco Congedo, Sandra Rousseau, and Christian Jutten. An Introduction to EEG Source Analysis with an Illustration of a Study on Error-Related Potentials. In Eduardo Reck Miranda and Julien Castet, editors, *Guide to Brain-Computer Music Interfacing*, pages 163–189. Springer London, January 2014.
- [5] M. Congedo, R. Phlypo, and Dinh-Tuan Pham. Approximate Joint Singular Value Decomposition of an Asymmetric Rectangular Matrix Set. *IEEE Transactions on Signal Processing*, 59(1):415–424, January 2011.
- [6] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [7] J. V Stone, J. Porritt, N. R. Porter, and I. D. Wilkinson. Spatiotemporal Independent Component Analysis of Event-Related fMRI Data Using Skewed Probability Density Functions. *NeuroImage*, 15(2):407–421, February 2002.
- [8] F.J. Theis, P. Gruber, I.R. Keck, A. Meyer-Base, and E.W. Lang. Spatiotemporal blind source separation using double-sided approximate joint diagonalization. In *Signal Processing Conference, 2005 13th European*, pages 1–4, September 2005.
- [9] Mohammad Niknazar, Hanna Becker, Bertrand Rivet, Christian Jutten, and Pierre Comon. Blind source separation of underdetermined mixtures of event-related sources. *Signal Processing*, 101:52–64, August 2014.
- [10] Xiao-Feng Gong, Xiu-Lin Wang, and Qiu-Hua Lin. Generalized Non-Orthogonal Joint Diagonalization With LU Decomposition and Successive Rotations. *IEEE Transactions on Signal Processing*, 63(5):1322–1334, March 2015.
- [11] V. Maurandi and E. Moreau. A Decoupled Jacobi-Like Algorithm for Non-Unitary Joint Diagonalization of Complex-Valued Matrices. *IEEE Signal Processing Letters*, 21(12):1453–1456, December 2014.
- [12] Marco Congedo, Louis Korczowski, Arnaud Delorme, and Fernando Lopes da Silva. Spatio-temporal common pattern: A companion method for ERP analysis in the time domain. *J. Neurosci. Methods*, 267:74–88, April 2016.
- [13] E. Moreau and O. Macchi. New self-adaptive algorithms for source separation based on contrast functions. In *IEEE Signal Processing Workshop on Higher-Order Statistics, 1993*, pages 215–219, 1993.
- [14] Louis Korczowski, Alexandre Barachant, Anton Andreev, Christian Jutten, and Marco Congedo. Brain Invaders 2 : an open source Plug & Play multi-user BCI videogame. In *6th International Brain-Computer Interface Meeting*, page 224, Pacific Grove, CA, United States, May 2016. BCI Society.

- [15] M. S. Bartlett. Smoothing Periodograms from Time-Series with Continuous Spectra. *Nature*, 161:686–687, May 1948.
- [16] John Polich. Updating P300: An Integrative Theory of P3a and P3b. *Clin Neurophysiol*, 118(10):2128–2148, October 2007.